Analysis of User Experience and Performance at Initial Exposure to Novel Keyboard Input Methods

Connor F. Scully-Allison, Hirav Parekh,
Frederick C. Harris Jr. and Sergiu M. Dascalu
Computer Science and Engineering, University of Nevada, Reno, USA
(cscully-allison, hparekh)@nevada.unr.edu, (dascalus, fred.harris)@cse.unr.edu

Abstract

We evaluated user performance and user experience with two novel input methods for mobile devices: Minuum and MessagEase. Subjects used a Qwerty keyboard to give a performance baseline. We compared input speeds, error rates, and keystroke counts among all three inputs to understand what factors discourage continued use or widespread adoption of new keyboard formats. It was found that MessagEase performed poorly upon initial exposure in terms of speed and error rate at 1482 mSec per character and 35.75% errors per line. Being 82.8% slower and 81.1% more error prone than Qwerty, there was a strong correlation between negative opinions of MessagEase and user performance. For Minuum, the performance gap was less significant at 32.1% slower speeds and 50.9% greater error rate. Accordingly Minuum was correlated with a better overall user experience compared to MessagEase.

keywords: Soft Controls; Empirical Study; HCI; Mobile User Experience; Mobile Keyboards; Touchscreen

1 Introduction

The final layout of the Qwerty keyboard was established on August 1, 1882. It was shipped with the Remmington Type-Writer No.2 and within a decade became the *de facto* standard key layout for mechanized text entry.[7] Over the following century, the typewriter became ubiquitous in office and research environments as the primary means of efficiently transcribing and storing data. From there, it seems almost inevitable that it would be adapted to new frontiers of computing technology and become the standard text input layout of the digital age.

Despite prodigious growth in technology, research, and innovation all pointing to better methods of text entry on a computer,[4] Qwerty remains entrenched as the default text input layout for nearly all human-computer interaction. With the growth of mobile computing technology, however, this entrenchment becomes more problematic by the year as our primary means of computer interaction becomes less analogous to using a typewriter and more analogous to finger painting. Accordingly, we propose a study to see if Qwerty can be unseated and what barriers stand in the way of such a technological coup d'etat.

It is well documented that technically superior, novel, approaches to computer interaction have a significant barrier of adoption when they are sufficiently different from extant standards.[2] As is the case with keyboard layouts Qwerty v. Dvorak, a quantifiably superior layout (Dvorak) fails to supplant an inferior layout (Qwerty) due to the investment of learning the new technique. [6] Unlike the case with Dvorak, however, modern soft controls offer a flexibility and ease of implementation which should make the transition from archaic interfaces to modern ones easy and affordable.[8]

We studied this phenomenon as it occurs with modern input software to understand what attracts potential users or turns them away. We gathered data on the quantifiable usability gap between a novel approach to mobile text input versus the accepted Qwerty layout at a user's first exposure (T_0). By asking qualitative questions to the user, we also measured how the initial gap feels to users and how it affects their future intention to explore novel approaches to computer input.

Even if a keyboard or other input method proves to be more efficient in the long run it does very little to unseat an established method if no one ever adopts it long term. So, in seeking to make a commercially successful keyboard application innovators must consider a user's response to their software at first exposure with equal or even greater significance than the long term. We hope that the data presented in this paper proves informative to innovators and researchers seeking to make a viable and marketable alternative to Qwerty.

2 Methodology

2.1 Participants

The participants of this experiment were men and women, between 18-30 years of age. We chose this group because they represent a demographic adaptable to shifting technology and most inclined to try out new software. To offset some possible skew due to convenience sampling we did recruit participants off campus as well as on, ending with a ratio of 3 off campus participants to 5 on.

2.2 Apparatus

The hardware used for this experiment was a Nexus 6p smartphone equipped with the three keyboards being measured: Android Standard Qwerty, MessagEase and Minuum. MessagEase was chosen for it's distinctness from the qwerty layout and its potential for very high input speeds, managing to claim a world record for input speeds in 2012.[5] By comparison, Minuum was chosen because of its similarity to the qwerty layout. We wished to study if a gap could be bridged between familiar and novel layouts and we thought Minuum represented a good hybrid layout.

The software used for logging data from the user was a custom web interface hosted on a free web server. The interface prompted the users to input static phrases into text input boxes, switch keyboards at appropriate times (with guidance from the experimenters) and pause to provide qualitative data.

On the back end, each character input by the user and the time between key presses in mSec was logged as an array of ordered pairs for each phrase entered. All arrays of ordered pairs were ultimately stored as JSON strings for later analysis.

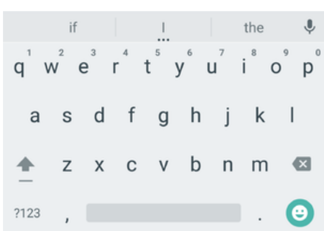

Figure 1: Standard Android M Keyboard

2.3 Procedure

The research procedure was highly iterative to derive as much information as possible from a limited scope of time. We met various participants at a University Library, walked them through the basic elements of our experiment, explained timelines, tasks, and user expectations.

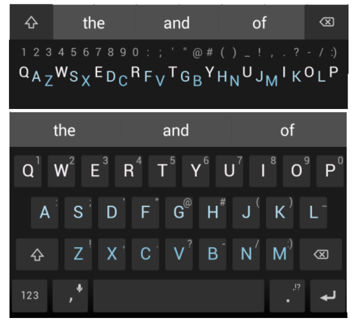

Figure 2: Minuum Mobile Keyboard App

Figure 3: MessageEase Mobile Keyboard App

We followed this initial introduction with a brief questionnaire gauging the subject's prior experience with novel keyboard layouts, preferences, experience texting or typing on a mobile keyboard, approximate estimations of daily use, etc. The questionnaire was comprised of several statements which participants evaluated on a Likert scale with 1 being "Strongly Disagree" and 7 being "Strongly Agree". An excerpt of the questionnaire follows:

1. I am familiar with mobile devices.
2. I frequently use a mobile keyboard daily.
3. I am interested in using a keyboard layout different from the standard Qwerty on my phone.

. . .

10. I would be interested in trying a new keyboard if I could type faster.

After this brief questionnaire we continued onto a Initial Exposure Phase. We exposed subjects to the keyboards, explained basic functionality and asked them to input "The Quick Brown Fox Jumped Over the Lazy Dog" 2 times with each. We then asked them several questions to gauge their feelings on the intuitiveness of the input method or gauge their visceral emotional responses to the keyboard. Another excerpt on the same likert scale as before:

1. I felt frustrated using the Minuum keyboard.
2. I felt frustrated using the MessagEase keyboard.

. . .

8. I would like to continue to use MessagEase as my primary keyboard in the future.
9. I would like to continue to use Minnum as my primary keyboard in the future.

From this we moved on to the principle testing phase. Here we asked subjects to input several atomic lines of text in a row. Participants were prompted to input their text on a standard keyboard, move to one novel keyboard and then switch to the other.

After this final testing phase, where we acquired our quantitative data, we gave the subjects one final survey to determine if a brief further exposure to these keyboards changed their opinions from their initial exposure. We also asked several open ended questions to get opinions on what participants liked most or least about each keyboard.

2.4 Tasks

Users were prompted to input four static lines of text with each keyboard. These lines were chosen arbitrarily off of Twitter:

1. With its blue skies and abundant life, planet Earth is a pretty special place.
2. A salmon's sense of smell is thousands of times better than a dog's.
3. When a book lover is in charge of the floor plan.
4. The net cooling effect of a healthy tree is equal to 10 air conditioners operating 20 hours a day.

Users were also instructed and frequently reminded to input text as quickly as possible. Subjects were additionally instructed to preserve all errors and attempt to minimize their use of the backspace key as much as possible.

2.5 Design

Our experiment was a single factor within-subjects design with 3 levels. The independent variable utilized in this experiment was the Keyboard subjects were instructed to use.

The total number of measurements was 8 participants x (3 sets of text x (293 input times in a set + 4 error rate per set + 4 keystroke count per set)) = 7224.

3 Experimental Results

After careful consideration, we determined that the most meaningful data for analysis would be a standard arithmetic mean of each dependant variable measured within subjects. The result of our calculations gave us the following data sets, rendered here tables and graphs, beginning with input speed:

3.1 Speeds

Table 1: Input Speeds Per Character

Avg. Per-Character Input Speed (mSec)			
	Android	MessagEase	Minuum
Subject 1	147.25	1609.75	260.75
Subject 2	284	1699.25	492.5
Subject 3	372.75	1475.5	467.75
Subject 4	299.5	1601.75	518
Subject 5	243.5	1240.75	307.25
Subject 6	294.75	2023.75	429
Subject 7	209.75	1400.25	297.25
Subject 8	192	806	237.5

Table 1 shows the average input speed per character for all the participants with three different keyboards. The mean input speed per character for the standard Android keyboard was 255.4 mSec. This was 82.8% lower than mean input with MessagEase at 1482.13 mSec and 32.1% lower than input with Minuum at 376.25 mSec. Between Minuum and MessagEase we observed a difference of 74.6% in input speeds, with MessagEase the slower of the two. The difference between all three of these input methods was statistically significant ($F_{2,14} = 105.3$, p <0.0001).

It is worth noting that the actual p-value computed was functionally zero, which is very uncommon. However, after an analyzing the distribution of data this does seem to be consistent with our measurements. Since p-value merely describes the probability that the variation in data was caused by random chance, it's evident, at a glance that the data appears consistently influenced by the non-random trials of the experiment. Figure 4. indicates the distribution of input speeds and shows that most of this data is very tightly organized around its own mean with small standard deviations.

We see in this data that there there is a high variance between input speeds per-character depending on the

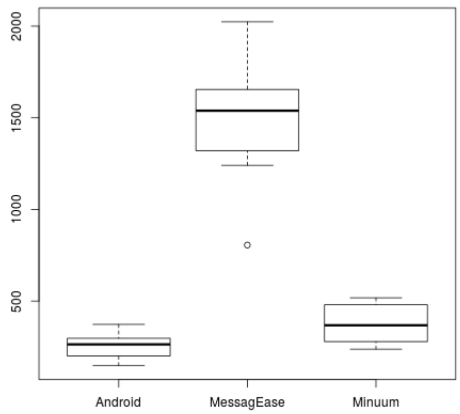

Figure 4: Box & Whisker Plot of Speed Averages

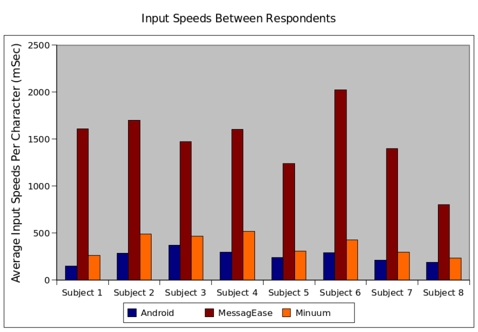

Figure 5: Average Input Speeds Per Character

keyboard used, however it is proportionally uniform across subjects. We also see strong evidence in support of the assumption that new interfaces are slower for input compared to learned ones. While, in of itself, this is not a particularly novel or interesting observation we can draw some conclusions from it in this instance.

Primarily we wish to comment on the magnitude of slowdown as subjects moved to more unfamiliar keyboard layouts. This shift is very evident in Fig. 4 While the shift from the Standard to Minuum resulted in minimal slowdown (about 100 mSec on average) the shift to MessagEase resulted in speeds about 6 times greater than standard. Although the makers of MessagEase promise to users that their software maximizes input speeds and enables users "to surpass . . . text entry speed with QWERTY" [3], an average consumer will never appreciate that speedup when she quits immediately because of such a severe perceived disadvantage. This vast difference in slowdown is almost certainly attributable to how similar Minuum is to a standard Qwerty vs. MessagEase which bears almost no similarity of layout.

The fact that input speed plays a strong influence on user experience is supported by our subject's collective responses to the statement "I would like to use MessagEase as my primary keyboard in the future." 5 of 8 respondents either disagreed or strongly disagreed with this statement indicating that a slow typing speed at T_0 is enough to push away a majority of users from continued use. By comparison, Minuum, with it's significantly smaller gap in typing speed produced generally a more even distribution of answers to the same question. 5 of the 8 respondents replied that they were either in favor of continuing to use Minuum or at least neutral to the idea. While it's anticipated that generally people will be resistant to change, people had less strong negative responses to the continued use of Minuum vis a vis MessagEase.

It should be noted as an interesting counterpoint that we did note a rapid improvement among user performance with MessagEase as we compared their four trials. Input speeds decreased from 2400ms per character, on average, down to 1325 ms by the end of the 4th phrase. This rapid improvement went unnoticed by subjects however as a majority (5 of 8 respondents) replied "Disagree" or "Strongly Disagree" to the statement "I feel as though I now type quickly with MessagEase."

3.2 Error Rates

Table 2: Average Error Rate Per Line

Avg. Error Rate Per Line (%)			
	Android	MessagEase	Minuum
Subject 1	18	41	22
Subject 2	0	30	11
Subject 3	13	26	9
Subject 4	0	41	4
Subject 5	2	31	24
Subject 6	2	52	7
Subject 7	10	24	18
Subject 8	9	41	15

Table 2 shows the mean error rate for all the participants with three different keyboards. The mean error rate for the standard Android keyboard was 6.75%. This was 81.1% lower than input with MessagEase at 35.75% and 50.9% lower than input with Minuum at 13.75%. Between Minuum and MessagEase we observed a difference of 61.5% in error rates, with MessagEase more error prone than Minuum. The difference between all three of these input methods was statistically significant ($F_{2,14} = 27.56$, p <0.0001).

The error rate metric chosen was one outlined in the

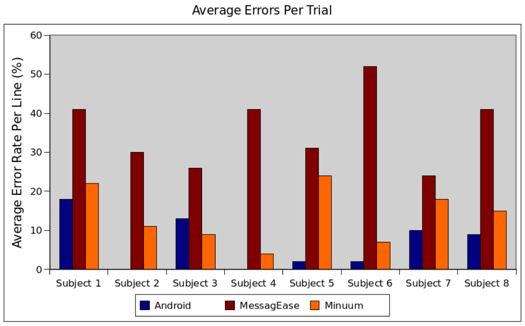

Figure 6: Average Error Rate Per Line

paper Analysis of Text Entry Performance Metrics[1], and is the most common approach to error rate. We evaluated it as the number of incorrect characters in the text over the total length of the transcribed text. As visible in the graph (Fig. 6), this produced a general map of high error rates with MessagEase, slightly fewer with Minuum and the least with the standard QWERTY keyboard.

The fact that error rates are lower with the standard keyboard appears to be consistent with expectations. Both speed and rates of error should be lowest in the input device which users are most familiar with. However, one aspect of this data which we find unintuitive is that MessagEase has the higher error rate on average compared to Minuum despite requiring much slower input. We expected that when people were forced to slow down they would be more careful in their work, however this data seems inconsistent with that. Our estimation is that error rate is likely lower in the case of Minuum due to the predictive text feature which drives it. It is likely that the computer aided suggestions and corrections provide more accurate input for any given word over an input type which offers no suggestions, in this case, MessagEase.

Comparing this against our qualitative data we find that high instances of error rates and slower speeds together have some correlation with user frustration. Among respondents those who had the highest error rates and slowest input speeds we saw that they responded strongly with “Agreement” or “Strong Agreement” to the statement “I felt frustrated using the MessagEase keyboard.” By contrast, these respondents, exhibiting faster speeds and lower errors with Minuum expressed “Disagreement” when asked the same of Minuum.

3.3 Keystrokes

Table 3 shows the mean keystroke count for all the participants with three different keyboards. The mean keystroke count for the standard Android keyboard

Table 3: Average Keystrokes Per Line

Avg. Keystrokes Per Line			
	Android	MessagEase	Minuum
Subject 1	110.75	73.75	171.5
Subject 2	112.25	61.5	170.5
Subject 3	80.25	75.5	148.75
Subject 4	114.25	76.75	154.25
Subject 5	112.25	71.75	144.25
Subject 6	109.5	59	99
Subject 7	113.75	73	142.5
Subject 8	109.75	75.25	144.5

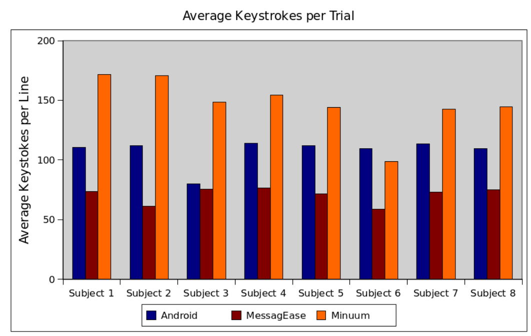

Figure 7: Average Keystroke Count Per Line

was 107.84. This was 52% higher than input with MessagEase at 70.81 and 26.6% lower than input with Minuum at 146.91. Between Minuum and MessagEase we observed a difference of 51.8% in keystroke count. The difference between all three of these input methods was statistically significant ($F_{2,14} = 55.604$, p <0.0001).

With keystrokes we see an inverse correlation to the previous two data sets. As can be seen in Fig. 7, the keystrokes for Minuum were highest amongst all keyboards while MessagEase claimed the lowest. The cause of this data inversion is unknown but we hypothesize that it is due to either the phenomenon of Minuum suggesting a whole collection of wrong words requiring users to start the word over again, users forgetting their place in a word since Minuum only populates suggestions and does not show a word grow as it is input, or a measurement error in the apparatus.

With most qualitative data favoring Minuum over MessagEase this seems to indicate that keystroke count is not a significant factor determining likability of a novel input method. We also speculate that the low keystroke count with MessagEase is likely a function of it’s low speed and would increase given greater fluency with this software. It would probably be approximately par with the android keyboard as users more clumsily enter text and delete it.

3.4 User Opinions

When asked about keyboard preference at the end of this experiment subjects responded that they preferred the Qwerty keyboard over MessagEase and Minuum. Despite this general rejection of these particular keyboards, it should be noted that 5 of 8 subjects replied either "Agree" or "Strongly Agree" to the statement "I would be interested in trying a new keyboard other than Minuum or MessagEase in the future." This response shows that even if current layouts fail to capture consumer interest there is still a market for novel keyboards in the 18-32 year-old demographic.

4 Conclusion

Ultimately, the results of this experimental study did support intended research goals. With a wide variance in input speeds, error rates and keystroke counts there was shown to be a significant difference with user performance between keyboards at T_0. Specifically, this study contributed three very important observations.

It produced strong quantitative data showing the gap in performance which a novel keyboard must overcome to match a standard method. It gave us extensive qualitative data which showed that this gap is not negligible and directly affects the users mood, experience, and intent to use a given method. And, finally, with a smaller performance and usability gap measured between Minuum and the standard keyboard we have a quantifiable model of what a more immediately successful keyboard application might look like.

We theorize that the most effective keyboard would likely be a progressive design that changes from qwerty to novel over time. Having established that the greatest barrier to adoption of novel keyboards is the steep learning curve, we imagine that a dynamic design could perhaps solve this problem by slowly introducing small novel aspects into a traditional qwerty until the keyboard until the keyboard becomes wholly novel.

5 Future Work

In terms of future work, we believe there is a great deal of room in HCI for studying what specific elements *will* produce an immediately user friendly novel keyboard. And, if such elements cannot be integrated into a UI, determine what external motivations might compel users to pick up these input methods and stay with them. We also see room for the continued research and development of a prime keyboard layout for mobile devices based on this data and other works.

6 Acknowledgement

This material is based upon work supported by the National Science Foundation under grant numbers IIA-1329469. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] A. S. Arif and W. Stuerzlinger. Analysis of text entry performance metrics. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pages 100–105, Sept 2009.

[2] Parmit K. Chilana, Mary P. Czerwinski, Tovi Grossman, Chris Harrison, Ranjitha Kumar, Tapan S. Parikh, and Shumin Zhai. Technology transfer of hci research innovations: Challenges and opportunities. In *Proc. of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 823–828, New York, NY, USA, 2015. ACM.

[3] Exideas. Message ease - the smartest touchscreen keyboard. `http://www.exideas.com/ME/index.php`. [Online; accessed 10-May-2016].

[4] I. Scott MacKenzie and Shawn X. Zhang. The design and evaluation of a high-performance soft keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 25–31, New York, NY, USA, 1999. ACM.

[5] MessagEase. New world record. `https://twitter.com/MessagEase/status/235412902423703552`, Aug 2012. [Online; accessed 11-January-2017].

[6] Jan Noyes. The qwerty keyboard: A review. *International Journal of Man-Machine Studies*, 18(3):265281, Mar 1983.

[7] Koichi Yasuoka and Makoto Yasuoka. On the prehistory of qwerty. *ZINBUN*, 42:161174, 2011.

[8] Shumin Zhai, Michael Hunter, and Barton A. Smith. The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, UIST '00, pages 119–128, New York, NY, USA, 2000. ACM.